Terpenes, Hormones, and Life: Isoprene Rule Revisited

Society for Endocrinology
JOURNAL OF ENDOCRINOLOGY

Terpenes, Hormones, and Life: Isoprene Rule Revisited

Journal:	*Journal of Endocrinology*
Manuscript ID	JOE-19-0084.R1
Manuscript Type:	Unsolicited Review
Date Submitted by the Author:	n/a
Complete List of Authors:	Hillier, Stephen; University of Edinburgh, MRC Centre for Reproductive Health Lathe, Richard; University of Edinburgh, Division of Infection and Pathway Medicine, University of Edinburgh Medical School
Keywords:	isoprene, terpene, steroid, evolution, Ruzicka

Title

Terpenes, Hormones, and Life: Isoprene Rule Revisited

Authors

Stephen G. Hillier[1] and Richard Lathe[2]

Addresses

[1]Medical Research Council Centre for Reproductive Health, University of Edinburgh, The Queen's Medical Research Institute, and [2]Division of Infection and Pathway Medicine, University of Edinburgh Medical School, Edinburgh, UK

Corresponding Author

Professor Stephen G. Hillier

Medical Research Council Centre for Reproductive Health, University of Edinburgh, The Queen's Medical Research Institute, 47 Little France Cresent, Edinburgh EH16 4TJ, UK

e-mail address: steve.hillier@ed.ac.uk

Short Title

Isoprene Rule Revisited

Key Words

isoprene; terpene; steroid; evolution, great oxidation event; Ruzicka

Word Count

5133

Abstract

2019 marks the 80th anniversary of the 1939 Nobel Prize in Chemistry awarded to Leopold Ruzicka (1887–1976) for work on higher terpene molecular structures, including the first chemical synthesis of male sex hormones. Arguably his crowning achievement was the 'biogenetic isoprene rule', which helped to unravel the complexities of terpenoid biosynthesis. The rule declares terpenoids to be enzymatically cyclized products of substrate alkene chains containing a characteristic number of linear, head-to-tail condensed, C_5 isoprene units. The number of repeat isoprene units dictates the type of terpene produced (i.e., 2, monoterpene; 3, sesquiterpene; 4, diterpene, etc.). In the case of triterpenes, six C_5 isoprene units combine into C_{30} squalene, which is cyclized into one of the signature carbon skeletons from which myriad downstream triterpenoid structures are derived, including sterols and steroids. Ruzicka also had a keen interest in the origin of life, but the pivotal role of terpenoids has generally been overshadowed by nucleobases, amino acids, and sugars. To redress the balance we provide a historical and evolutionary perspective. We address the potential abiotic generation of isoprene, the crucial role that polyprene terpenoids played in early membranes and cellular life, and emphasize that endocrinology from microbes to plants and vertebrates is firmly grounded on Ruzicka's pivotal insights into the structure and function of terpenes. A harmonizing feature is that all known lifeforms (including bacteria) biosynthesize triterpenoid substances that are essential for cellular membrane formation and function, from which signaling molecules such as steroid hormones and cognate receptors are likely to have evolved.

Introduction

> The structural similarities of the higher terpenes raise the question as to whether these compounds may have been formed according to a uniform principle in nature (Ruzicka 1966).

Terpenes (including sterols, steroids, and related aromatic hydrocarbons) are present in all known life forms where they pivotally impact individual and population survival (Summons *et al.* 2006; Nes 2011; Jiang *et al.* 2016). Fossilized terpenes have been discovered in geological deposits billions of years old, signifying involvement in the very beginnings of life on Earth (Ourisson & Albrecht 1992; Melendez *et al.* 2013): involvement so fundamental that as ancient lipids they may represent (along with DNA, RNA, and protein according to the central dogma), a fourth molecular strand of terrestrial life.

The terpenome (the compendium of all known terpenoids) is so vast that it 'accounts for nearly one-third of all compounds currently characterized in the *Dictionary of Natural Products*' (http://dnp.chemnetbase.com)' (Christianson 2017). Why so many structurally related compounds with such diverse functions should exist has yet to be explained. The clue seemingly lies in their core chemical structure: they are all composed of multiples of a 5-carbon unit called isoprene (2-methyl 1,3-butadiene, isopentene), one of the most common organic chemicals on Earth (Sharkey & Yeh 2001).

2019 marks the eightieth anniversary of the 1939 Nobel Prize in Chemistry awarded to Leopold Ruzicka (Ružička[1]), a Croatian national who worked extensively in Switzerland, for research that established the importance of the 'isoprene rule' in the elucidation of

[1] The pronunciation can be best explained by the French transcription 'Rougitchka' (https://www.nobelprize.org/prizes/chemistry/1939/ruzicka/biographical/).

terpene chemical structures, including classification of cholesterol and sex steroids as triterpenoids. He also advanced a biochemical extension of the rule that became the basis of present day understanding of terpene biosynthesis, known as the 'biogenetic isoprene rule' (Ruzicka 1953; Ruzicka 1959).

The fundamental significance of the biogenetic isoprene rule to biology, immediately evident at the time, continues to become increasingly apparent. In the same way as terpene chemistry is built on a common molecular plan involving isoprene, so too is terpenoid biology.

Ruzicka became increasingly interested in the possible link between the biogenetic isoprene rule and the question of life's origin (Eschenmoser 1990). The occasion of his 80th Nobel anniversary allows the opportunity to celebrate his seminal contributions and reappraise the connection between terpenoid biochemistry and the existence of life on Earth.

The Isoprene Rule

> […] the leading question was whether the carbon skeletons of the higher terpenes were also composed of isoprene units. For all the compounds we examined, the answer was positive, and thus the original working hypothesis gradually grew into the isoprene rule (Ruzicka 1959).

Isoprene, the prototypic terpene substance, is one of the most copiously produced volatile hydrocarbon chemicals on Earth (McGenity *et al.* 2018) owing to the global abundance of terpenoid biosynthesis, not vice versa. Approximately 40% of the biogenic volatile organic

compounds emitted by plants are in the form of isoprene, and isoprene is the principal hydrocarbon identified in human breath (Gelmont *et al.* 1981). Cinema audiences exhale more isoprene when watching scenes of suspense, and isoprene levels spike in the air above football fans when goals are scored (Stonner & Williams 2016).

The centrality of isoprene to terpene chemistry cannot be overestimated (Nes 2011). First obtained from burning rubber by Michael Faraday (Faraday 1826), isopentene was isolated as a distillation product of natural rubber by Greville Williams (Williams 1860) who correctly assigned it the empirical formula of C_5H_8 and named it isoprene. Tilden (Tilden 1884) went on to distil isoprene from turpentine oil and showed that it could be dimerized into 'dipentene' ($C_{10}H_{16}$) by heat or treatment with sulfuric acid, portending modern discussions on the abiogenesis of isoprenoids (see below). Meanwhile, Kekulé coined the name *terpen* (English 'terpene') for the group of hydrocarbons obtained from turpentine oil with C:H ratios of C_{10}–C_{16}, of which isoprene (dipentene) was one (Box 1). Remarkably, all this was accomplished before isoprene's iconic molecular structure was formally established (Ipatiew & Wittorf 1897) (Figure 1).

When Otto Wallach began his systematic studies of the terpenes, he observed that individual terpene structures contained multiples of a 5-carbon unit that allowed them to be classified according to the number of such units they contained (Wallach 1887). The canonical C_5 unit proved to be isoprene, which Wallach recognized as the core terpenoid structure. That terpenes might be represented as repeating isoprene units became known as the 'isoprene rule'. In this scenario, C_{10} monoterpenes contain two head-to-tail linked hemiterpene isoprene units, C_{15} sesquiterpenes contain three C_5 isoprenes, etc. (Box 1). The isoprene rule was extended to accommodate C_{20} diterpenes ($4 \times C_5 = C_{20}$), C_{30} triterpenes ($6 \times C_5 = C_{30}$), and beyond (Ruzicka 1953; Eschenmoser 1990). It is

noteworthy that Wallach's rule was successfully applied a full decade before the molecular structure of isoprene had been fully identified (Ipatiew & Wittorg 1897) or its chemical synthesis unambiguously achieved (Euler 1897).

The isoprene rule languished until the 1920s when it was referenced during the structural determination of cholesterol (Wieland 1966; Robinson 1932) and shaped by Ruzicka into the 'biogenetic isoprene rule' (Ruzicka 1953; Ruzicka *et al.* 1953). Ruzicka's version of the rule allowed terpenoid structures to be explained or predicted based on accepted reaction mechanisms involving acyclic precursors that were products of isoprene condensation, such as geraniol, farnesol, and geranylgerianol (Ruzicka 1963). His Nobel lecture vividly illustrates the success of this approach, which led to the classification of cholesterol as a triterpenoid substance several years before any formal demonstration that the carbon atoms in its side-chain follow the isoprene rule (Wüersch *et al.* 1952; Ruzicka 1966).

The rise of biochemistry in the 1940s and 50s saw the isoprene rule shaped beyond a chemical hypothesis for predicting terpenoid structure into a tool for delineating pathways of terpenoid biosynthesis. Within less than a decade, squalene was installed as an essential intermediate in sterol biosynthesis (Langdon & Bloch 1953), acetic acid was identified as the primary carbon source of squalene and sterol biosynthesis (Bloch & Rittenberg 1942; Langdon & Bloch 1953), mevalonic acid (MVA) (Tavormina *et al.* 1956; Amdur *et al.* 1957) and farnesyl pyrophosphate (FPP) were established as vital squalene and cholesterol precursors (Cornforth *et al.* 1958; Lynen *et al.* 1958), squalene cyclization was mapped (Woodward & Bloch 1953), and, above all, isopentenyl pyrophosphate (IPP) – the biologically activated form of isoprene – was identified as the founder C_5 unit upon which all terpenoid *biosynthesis* depends (Lynen *et al.* 1958). The tortuous route to unraveling these key pathways is aptly summarized by Bloch (Bloch 1992).

All natural triterpenoid substances could now be understood as end-products of a time-honored biogenetic sequence beginning with the formation of IPP – the active C_5 unit in terpenoid biosynthesis. Linear, head-to-tail, coupling of IPP with its isomer, dimethylallyl pyrophosphate (DMAPP), and a secondary molecule of IPP leads to C_{15} FPP via C_{10} geranyl pyrophosphate (GPP) (Figure 2). Reductive, tail-to-tail dimerization of two FPPs leads to formation of C_{30} squalene (Popjak *et al.* 1969). Importantly, the biochemical route from IPP to squalene is shared by all three domains of life, consistent with the ubiquity of terrestrial terpene biology.

Beyond Squalene

> […] a scheme has been developed leading from squalene to the formulae of the basic representatives of all known cyclic triterpene groups […]. This result is considered to support the squalene hypothesis of the biogenesis of cyclic triterpenes (Eschenmoser *et al.* 1955).

The isoprene rule morphed into the 'squalene rule' when squalene proved to be an essential intermediate in the biosynthesis of cholesterol and the carbon skeletons of all other then known cyclic triterpenes (Eschenmoser *et al.* 1955). The paper positing cyclization of squalene as the axis of triterpenoid biogenesis (Woodward & Bloch 1953) was published in the same year that Watson and Crick disclosed the structure of DNA (Watson & Crick 1953). Both discoveries had immediate impact and are still absolutely relevant to our understanding of the biochemical basis of life on Earth. Fittingly, all four authors duly received Nobel prizes. (Watson and Crick shared the Physiology or Medicine

prize with Maurice Wilkins in 1962; Bloch shared the 1964 Physiology or Medicine prize with Feodor Lynen, and Woodward was awarded the full Chemistry prize in 1965.)

Squalene cyclization proceeds as an electrophilic reaction cascade catalyzed by phylum-specific terpene cyclase enzymes – whereby the C_{30} polyalkene chain is effectively rolled into a polycyclic structure comprising up to five interconnected carbon rings and a residual side-chain of variable length. Depending on biological context, squalene cyclization occurs with absolute regio- and stereospecificity to produce the signature multicyclic carbon skeletons from which myriad downstream terpenoid structures are derived (Nes 2011; Xu *et al.* 2004; Jia & Peters 2017).

In bacteria, squalene cyclization is catalyzed by squalene cyclase and does not require prior epoxidation (Summons *et al.* 2006) (Figure 3). By contrast, in eukaryotes squalene cyclization catalyzed by oxidosqualene cyclase is preceded by 2,3-epoxidation. This is the first oxygen-dependent step in steroid formation, and provides the point at which steroid biosynthesis in metazoans diverges from hopanoid biosynthesis in bacteria.

Ruzicka clearly saw the broader implications of the 'squalene hypothesis' in that it logically explained how such a constellation of multicyclic terpenoid structures could be derived from a single precursor molecule (Figure 4) (Ruzicka 1959).

Polycyclic Terpenoids

> After Kekulé, in 1865, first introduced the carbon ring into structural chemistry in his formula for benzene, the 6-membered ring maintained its unique position in the taxonomy of organic chemistry for several decades (Ruzicka 1966).

The molecular mechanism of squalene cyclization to lanosterol, the primary cholesterol precursor, was established within a decade of Ruzicka's 1945 Nobel Lecture (Woodward & Bloch 1953; Eschenmoser *et al.* 1955; Eschenmoser & Arigoni 2005) (Figure 5). Lanosterol is now recognized as the major terpenoid precursor for fungal as well as animal sterols and steroids, whereas cycloartenol gives rise to β-sitosterol and downstream sterol and steroidal metabolites in plants with a photosynthetic lineage (Nes 2011).

Extant animal, plant, and fungal clades have terpenoid signatures corresponding to individual needs for survival within particular ecosystems. Of the estimated >150 multicyclic carbon skeletons known or hypothesized to serve as triterpenoid precursors, only one (lanosteryl cation) is the source of cholesterol and true steroids. The rest are the uniquely adapted sources of the colors, scents, poisons, potions, rubbers, and waxes that are, worldwide, the components of the terpenome (Xu *et al.* 2004; Jiang *et al.* 2016)

Terpenoids based on the tetracyclic 6-6-6-5 lanostane carbon skeleton form a subsection of the terpenome known as the sterolome (Nes 2011). The sterolome is estimated to comprise at least 1000 natural products derived from lanosterol and related molecules that carry out essential biological functions across the three domains of life on earth. Cholesterol is the parent animal sterol, while cycloartenol and ergosterol are major plant and fungal equivalents. Chemically, they differ mainly in their side-chain substitution and saturation, as well as in their degree of esterification into glucuronides, glycosides, sulfates, etc. In animals cholesterol is famously converted to bile acids, vitamin D, and true steroids. The major plant sterols are ethylated (β-sitosterol, stigmasterol) or methylated (cycloartenol) at C_{24} of the side-chain, and the balance between ethylation and methylation is specific for individual plant species (Valitova *et al.* 2016). Plants also metabolize cholesterol into brassinosteroids (plant steroids), glycoalkaloids, cardenolides (poisons),

saponins (vegetable soaps), and withanolides (plant defense substances). Present-day insects do not biosynthesize cholesterol *de novo* but metabolize dietary molecules such as cholesterol into ecdysteroids (moulting hormones). Nematodes convert cholesterol to worm-specific pheromonal sterols (dafachronic acids) (Aguilaniu *et al.* 2016). Fungal steroids produced from ergosterol include antheriodiol (female sex hormone) and oogoniol (male sex hormone) (Nes 2011).

Thus, although steroids can be regarded as specialized higher terpenoids that fulfill multiple functions vital to metazoan life, the entire sterolome occupies only a miniscule corner of global terpenoid biochemistry.

Special mention must be made of the pentacyclic 6-6-6-6-5 hopanoids. First identified as a terpenoid component of resin extracted from plant genus *Hopea*, they exist in diverse bacteria and certain lower plant forms such as algae and lichens, where they function as membrane lipids (Saenz *et al.* 2015; Belin *et al.* 2018). The structural and functional similarity of bacterial hopanes to eukaryotic cholestanes renders them of particular interest in interpreting the molecular beginnings of life, not least because geological evidence suggests that hopanoid sterols were (and perhaps still are) among the most abundant natural products on Earth (Ourisson & Albrecht 1992).

Ancient Organics

Geologically durable isoprenoids have existed on Earth since the dawn of life. Hydrocarbons assembled from repeating isoprene units are ubiquitous in ancient sediments. Biomarker evidence for eukaryotes comes from terpenoid steranes with diagnostic alkylation patterns in Barney Creek Formation rocks 1.64 billion years (Ga)

before present. ((Summons *et al.* 1988), reviewed in (Brocks & Summons 2003)). Steroid residues have been found in rocks dated to ~2.7 Ga (Summons *et al.* 2006) and cholestanes have been identified in Ediacaran (protoist) fossils, which are the oldest confirmed macroscopic animals in the rock record (Bobrovskiy *et al.* 2018). The hopane carbon skeleton is particularly ubiquitous in oil shales and petroleum deposits in sedimentary rocks. Above all else, the presence of hopane residues in geological deposits laid down before the advent of atmospheric oxygen (see below) implies a fundamental contribution of sterols to the earliest stages of the evolution of terrestrial life. However, as discussed in the next section, the origins of prebiotic terpenoids remain enigmatic.

Isoprenoid Abiogenesis

> […] the level of perfection achieved by organic chemistry […] is enabling biochemistry to penetrate the innermost secrets of life processes on a molecular basis. Attributed to Ruzicka (Eschenmoser 1990).

Simple precursor molecules in the primeval soup fuelled the origin of life (Haldane 1929; Oparin 1953); these can be generated from a prebiotic milieu containing only H_2O, CH_4, NH_3, and H_2 (Miller & Urey 1959). Sugars are polymerization products of HCHO, amino acids are generated by condensation of HCN, HCHO, and NH_3, and bases are polymers of HCN/NH_3 and/or formamide. Reactive phosphorus moieties were no doubt abundant (Yamagata *et al.* 1991; Schwartz 2006; Pasek *et al.* 2013). Thus, precursors to the first three strands of life according to the central dogma (DNA makes RNA makes protein) were available for molecular tinkering.

Much less attention has been paid to the 'fourth strand' – lipids – that are essential for the generation of the first micelles and coacervates, the inferred precursors to cellular life (Segré & Lancet 2000). Indeed, the 'lipid world' undoubtedly accompanied, or even preceded (Ourisson & Nakatani 1994; Segré *et al.* 2001), the emergence of nucleic acid-based life.

Abiotic synthesis of straight-chain hydrocarbons (up to pentane) in Urey–Miller experiments has been confirmed, and extraterrestrial pentacarbon molecules have been widely detected (McGuire 2018). On Titan, the major moon of Saturn, lakes of methane and ethane are present, but these are not the final end-products, and complex photochemical conversions generate diverse hydrocarbons at high altitudes in the Titan stratosphere (Wilson & Atreya 2018; Waite, Jr. *et al.* 2007). The terpenoids pristane and phytane have been reported in samples of meteorites falling to Earth (Oro *et al.* 1966), and Cronin and Pizzarello (1990) found C_{15} to C_{30} branched alkyl-substituted mono-, di-, and tricyclic alkanes in the Murchison meteorite (Cronin & Pizzarello 1990) that fell in Australia in 1969. While it has not been formally possible to exclude terrestrial contamination (reviewed in (Sephton 2002)), the scene is set for long-chain hydrocarbons, both saturated and unsaturated – although the inferred preponderance of unsaturated C_5 isoprene precursors remains unexplained.

Following Ruzicka, isoprene was a likely prebiotic precursor for polymerization into linear polymers (Lazcano *et al.* 1983), but the abiotic origin of isoprene remains uncertain. Because isoprene is normally a gas (boiling point = 34.1°C), and is intensely insoluble in water, early Miller–Urey-type experiments may have overlooked this important avenue. In his sketches, Ruzicka (reproduced in (Eschenmoser 1990), p 9) outlined the formation of 2-butenal (CH_3–CH=CH–CHO, a potential isoprene precursor) from elementary

components, and no doubt intended to take this further. Ourisson and colleagues (Nakatani *et al.* 2012) suggested that isoprene could have been formed from isobutene, ethylene, and formaldehyde at high temperatures (with further product being generated by condensation on clays), noting that both isoprene and HCHO are components of volcanic gases (Dong *et al.* 1994). Other potential routes involve acetic acid and acetylene, or deamidation and decarboxylation of leucine, an amino acid detected in Miller–Urey experiments in the presence of H_2S (Parker *et al.* 2011). However, none of these routes so far resoundingly explains the inferred preponderance of isoprene moieties in the prebiotic chemosphere. Given the vital importance of isoprenes to the emergence of cellular life, resolving the puzzle of isoprene abiogenesis and its inferred pivotal role remains a priority.

Differential solubility provides an insight. Polyisoprene synthesis requires activation of isoprene by (pyro)phosphorylation (e.g., through the generation of isopentenyl pyrophosphate, IPP, and its isomer, dimethylallyl pyrophosphate, DMAPP). Reactive polyphosphates are generously emitted by volcanic activity (Yamagata *et al.* 1991), and could react in the gas phase with C=C and C≡C hydrocarbons. Crucially, (unlike isoprene) pyrophosphorylated derivatives of isoprene (i.e., IPP and DMAPP) are relatively soluble in water (7 and 25 g/L, respectively). Dissolution would also have afforded protection from photochemical degradation. This combination of factors could have led to accumulation of isoprene (pyro)phosphates in prebiotic oceans, paving the way to later exploitation at the origin of life.

Given (pyro)phosphate derivatives of isoprene, the generation of long-chain isoprenoids is chemically unchallenging, but the cyclization of squalene to lanosterol and related molecules also represents an enigma because of the multiple different boat/chair configurations of the ring systems, whereas only a specific subset (or small number of

subsets) is seen in extant molecules of this class. In a cyclic dry-down and dilute scenario (e.g., (Lathe 2005)), it is possible that clays could have fostered specific configurations (see e.g., (Ourisson & Nakatani 1994)), and templated cyclization mediated by product (noting that lanosterol readily forms crystals (Liu & Sawant 2002)) is a further possibility.

Life and the Great Oxygenation Event (GOE)

> Somewhere in between (bio)chemical and biological evolution we must assume a point where life was created. Attributed to Ruzicka (Eschenmoser 1990).

At the origin of life the planetary atmosphere contained little if any oxygen. Any proto-lifeform must have been independent of oxygen, with its lipid biochemistry adjusted accordingly. By contrast, the synthesis of 'modern' long-chain fatty acids (LCFAs) and sterols (the major components of cellular membranes in all organisms except Archaea) requires oxygen for synthesis. This has important implications for the biosynthesis of prebiotic/co-biotic polymers.

Free O_2 only became available with the advent of oxygenic photosynthesis by primitive cyanobacteria-like organisms, leading to the onset of the GOE, where O_2 was an incidental (and toxic) byproduct of sequestration of carbon from CO_2 (Figure 6). The timing of this transition has been accurately dated, based on isotope ratios and paleomagnetic studies, to between 2.46 and 2.43 Ga (Lyons *et al.* 2014; Gumsley *et al.* 2017).

Low levels of free O_2 may have been available a little earlier (e.g., at 2.5–2.95 Ga (Anbar *et al.* 2007; Planavsky *et al.* 2014)), but solubility is a further factor constraining O_2 availability. Following the formation of the Earth–Moon system at around 5.5 Ga, the first

rains fell to give oceans with a temperature of ~100°C, and the mean temperature declined roughly linearly from that time until today ((Sleep 2010; Garcia *et al.* 2017), but see (Pope *et al.* 2012)) pointing to surprisingly high temperatures of ~90°C at the origin of life (ca 3.9 Ga, or possibly a little earlier (Battistuzzi *et al.* 2004); Figure 6), where O_2 was many-fold less soluble in water, further accentuating oxygen limitation.

Thus, life is presumed to have emerged, in the near-total absence of oxygen, through extreme thermophiles most closely related to the Archaea (Gribaldo & Brochier-Armanet 2006; Eme *et al.* 2017). Although the exact relationship between the Archaea, Bacteria, and Eukaryota remains contentious, for simplicity we retain the Archaea/Bacteria distinction, and follow the idea that Archaea preceded both Bacteria and Eukaryota, with the unusual membrane composition of Archaea (see below) representing an obstacle to alternative phylogenies (Eme *et al.* 2017).

Centrally, polyisoprenes can be synthesized in the absence of oxygen. It has been argued that terpenes (and not LCFAs) were the essential components of primitive membranes (Ourisson 1989; Ourisson & Nakatani 1994; Nakatani *et al.* 2012; Nakatani *et al.* 2014), as well as those of present-day Archaea (Langworthy *et al.* 1982). Indeed, polyprenyl phosphates spontaneously form membrane vesicles (Nakatani *et al.* 2014). That these molecules are the functional equivalents of LCFAs was elegantly demonstrated by the engineering of *E. coli* whose membranes contain up to 30% of archaeal lipids – these bacteria are viable, and under some conditions the hybrid membrane even confers a growth advantage (Caforio *et al.* 2018).

As with present-day membrane lipids, these early polyisoprene units are presumed to have been linked to glycerol. Glycerol phosphates may have been abundant at the origin

of life (Pasek *et al.* 2013); indeed, the triose glycerol may even have preceded the pentoses ribose/deoxyribose in nucleic acids (Joyce *et al.* 1987)). Importantly, however, the glycerol linkage of membrane terpenoids differs significantly from that of present-day LCFAs. Membrane polyisoprene units in early life (as in present-day Archaea) are inferred to have been linked to glycerol via ether bonds [R–O–R] (Langworthy *et al.* 1982; de Rosa M. *et al.* 1986; Sprott 1992), contrasting with the ester bonds [R–(C=O)–R] that typify modern membrane LCFA diacylglycerols, thereby further reducing the requirement for oxygen (Figure 7).

Sterol–Hopanoid Homology: Membrane Stabilization

> It has taken all my life's work to convince myself that life is chemistry; and now you come along and tell us it is physics (Ruzicka's reaction to Manfred Eigen's theory of evolution based on mathematical principles, as recounted by Eschenmoser (Eschenmoser 1990)).

Terpenoid-based membranes are compatible with life, but terpene-only membranes tend to be unstable. Cholesterol, by contrast, is a molecule beautifully crafted by evolution to stabilize the membrane bilayer. However, cholesterol synthesis requires multiple oxygen-dependent events (Bae *et al.* 1999) (Figure 3), and cholesterol synthesis could therefore not have taken place before GOE ((Summons *et al.* 2006); discussed in (Galea & Brown 2009)). Instead, in the absence of oxygen, squalene and other long-chain terpenes may have contributed to membrane structure – squalene can partially rigidify membranes (Spanova *et al.* 2012; Gilmore *et al.* 2013) and may have played a similar role in the earliest life forms.

Cyclization of squalene generates lanosterol and related molecules (Figure 5), such as bacterial hopanoids, that surpass squalene in their ability to rigidify membranes. All these molecules contain the 4-dimethyl motif of bacterial hopanoids and, given their hydrophobic nature and rigid planar structure, readily intercalate into lipid membranes where they exert significant stabilizing effects as functional equivalents of cholesterol ((Rohmer *et al.* 1979)); reviewed by Ourisson (Ourisson *et al.* 1987)).

Although squalene cyclization in eukaryotes can be precipitated by the addition of a single oxygen atom (Figure 3), oxygen-independent enzymatic cyclization of squalene could have taken place in bacteria via the squalene-hopene cyclase enzyme ((Reinert *et al.* 2004); discussed in (Poralla 2004)), the antecedent to prokaryotic and eukaryotic sterol synthases (Pearson *et al.* 2003). Thus, early bacterial and proto-eukaryotic membranes are likely to have contained, in addition to long-chain polyprene lipids, sterols related to lanosterol/hopanoids, whereas the emergence of modern-day cholesterols (and steroids) awaited the advent of plentiful atmospheric oxygen.

Terpenoids and the Origins of Steroid Signaling

> The ability of so many living organisms to biosynthesize all these compounds with always the same configuration, out of hundreds of alternatives, must surely be one of the most remarkable chemical achievements of Nature (Ruzicka 1956).

There is consensus that eukaryotes evolved via intracellular symbiosis (see above) probably between an Archaea (the host) and a bacterium (the endosymbiote), although Archaea and Bacteria may have lived in close association for a protracted period – only

later culminating in formal endosymbiosis (reviewed by (Gribaldo & Brochier-Armanet 2006; Martin *et al.* 2015)) (Figure 6). Given that bacteria are likely to have been capable of synthesizing lanosterol (a better membrane-stabilizer than squalene), whereas the Archaea lack the necessary enzymes (Desmond & Gribaldo 2009), this might have provided another driving force for symbiosis leading to the generation of eukaryotes (the first geological fossils of eukaryotes date to somewhere between 1.6 Ga and 2.1–2.3 Ga (e.g., (Butterfield 2015)). In this conceptual scenario, Archaea might have provided a soluble precursor (e.g., farnesyl pyrophosphate, FPP) and received lanosterol or similar molecule in return. This conjecture – that a triterpene was the currency of exchange – has implications for the evolution of nuclear receptors (NRs) and steroid signaling.

In higher eukaryotes, derivatives of terpenoid sterols and steroids orchestrate multiple aspects of growth, development, and reproduction by targeting intracellular NRs, but the origin of steroid signaling remains open to debate. The ancestral ligand for the first NRs may have been a terpenoid or long-chain fatty acid (Moore 1990; Bridgham *et al.* 2010). Studies of the sponge *Amphimedon queenslandia* that contains only two NR polypeptides revealed that both NR1 and NR2 bind long-chain fatty acids (Bridgham *et al.* 2010), but terpenoids were not investigated.

Were farnesoids the forerunners to steroids? In multiple species the sesquiterpene farnesol and its derivatives are the key signaling molecules, not steroids. The key quorum-sensing molecules in the yeast *Candida albicans* (Polke *et al.* 2018) and the juvenile hormones of insects and crustacea (Qu *et al.* 2018) are all farnesoids. In plants the cyclic farnesoid, abscisic acid, is a crucial phytohormone (Cutler *et al.* 2010), and the same molecule has also been implicated in signaling in vertebrates (e.g., (Bassaganya-Riera *et al.* 2011)).

Intriguingly, the terpenoid FPP (but not farnesol) can activate multiple present-day steroid NRs, and it was suggested that this might reflect a common structural feature that was present in an ancestral NR (Das *et al.* 2007) – with the inference that FPP was the primeval ligand for this class of receptors. In support, activation of diverse NRs by FPP has been confirmed (Goyanka *et al.* 2010) and FPP is to this day an important human metabolite that regulates oxidative stress, in part by acting through the glucocorticoid receptor (Pastar *et al.* 2016).

There are, moreover, intriguing sequence and structural similarities between extant squalene synthases (SQSs) – that assemble two molecules of FPP into a single squalene molecule – and the ligand-binding domains of NRs (R Lathe & and S G Hillier, unpublished observations). Perhaps SQS adopted an early signaling role in response to binding of its substrate, FPP? – pointing to terpenoid forerunners of NR signaling in the bacterial symbionts of the earliest eukaryotes. Further research to address this intriguing possibility is certainly warranted, and if confirmed this would cast light on the later emergence of steroid signaling in higher eukaryotes.

Conclusions

> Attempts may be made to interpret the isoprene rule, not only as a working hypothesis in the laboratory, but also as a structural principle employed by nature (Ruzicka 1966).

The isoprene rule was successfully used to elucidate the chemical structures of some of the most important natural substances on Earth. Ruzicka's special contribution was to

recognize the power of the rule to explain and predict complex polycyclic chemical structures. Crucially, he saw how the isoprene rule might provide a unifying principle for resolving the multiple mysteries of triterpenoid biochemistry. In the same way as Kekulé peered into the fire and saw the licking flames shape a 6-membered benzene ring, Ruzicka's musings on aromatic chemistry led to the multimembered alicyclic picture of the terpenoid world.

Ruzicka operated in an era when organic chemistry was informing the emerging discipline of bio(logical) chemistry. He is quoted by his longstanding colleague Albert Eschenmoser (Eschenmoser 1990) as saying 'To understand biochemistry you need to know at least as much organic chemistry as for organic chemistry itself'. The extrapolation from acid-catalyzed polyene cyclization (encountered in his early days as a perfume chemist) to enzymatically catalyzed squalene cyclization in natural triterpenoid biogenesis illustrates this truism.

He was particularly fascinated by the chemical origins of life on Earth and saw a potential role for isoprenoid chemistry in its understanding. His prescience is underscored by the evidence, which continues to grow, that unicellular and metazoan cell membranes share structural and functional similarities based on terpenoid biochemistry, pointing to lipids as a fourth strand in the evolution of life alongside nucleobases, sugars, and proteins.

Ruzicka's stereochemical vision was famously matched by his prowess as a synthetic chemist. Having classified cholesterol as a triterpenoid he entertained the possibility that sex hormones 'of the oestrane and androstane type' might be molecules (later known as steroids) in which the side-chain of cholesterol had been split off. He formally proved this to be the case, and thereby achieved the first synthesis of a sex hormone: androsterone.

This in turn led to his synthesis of the principle male sex steroid testosterone, for which he shared the 1939 Nobel Prize in Chemistry with Adolf Butenandt.

Author Contribution Statement

Both co-authors contributed equally to the manuscript.

Declaration of Interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the work presented.

Acknowledgments

The authors gratefully acknowledge Professors Albert Eschenmoser (ETH Zurich Laboratorium für Organische Chemie), Yoichi Nakatani (Université de Strasbourg), and J. Ian Mason (University of Edinburgh) for their helpful comments during preparation of the manuscript.

Funding

This research did not receive any specific grant from any funding agency in the public, commercial or not-for-profit sector.

References

Aguilaniu H, Fabrizio P & Witting M 2016 The role of dafachronic acid signaling in development and longevity in Caenorhabditis elegans: digging deeper using cutting-edge analytical chemistry. *Frontiers in Endocrinology)* **7** 12.

Amdur RH, Rilling H & Bloch K 1957 The enzymatic conversion of mevalonic acid to squalene. *Journal of the American Chemical Society* **79** 2646-2647.

Anbar AD, Duan Y, Lyons TW, Arnold GL, Kendall B, Creaser RA, Kaufman AJ, Gordon GW, Scott C, Garvin J & Buick R 2007 A whiff of oxygen before the great oxidation event? *Science* **317** 1903-1906.

Bae SH, Lee JN, Fitzky BU, Seong J & Paik YK 1999 Cholesterol biosynthesis from lanosterol. Molecular cloning, tissue distribution, expression, chromosomal localization, and regulation of rat 7-dehydrocholesterol reductase, a Smith-Lemli-Opitz syndrome-related protein. *Journal of Biological Chemistry* **274** 14624-14631.

Barton GA 1906 Tree worship. In *The Jewish Encyclopedia*, pp 269-270. Eds I Singer & et al. Funk and Wagnalls.

Bassaganya-Riera J, Guri AJ, Lu P, Climent M, Carbo A, Sobral BW, Horne WT, Lewis SN, Bevan DR & Hontecillas R 2011 Abscisic acid regulates inflammation via ligand-binding domain-independent activation of peroxisome proliferator-activated receptor gamma. *Journal of Biological Chemistry* **286** 2504-2516.

Battistuzzi FU, Feijao A & Hedges SB 2004 A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *BMC Evolutionary Biology* **4** 44.

Beckmann S 2012 Resin and ritual purification: terebinth in Eastern Mediterranean Bronze Age culture. In *Athanasia: The Early, the Celestial and the Underworld in the Mediterranean from the Late Bronze and the Early Iron Age*, pp 27-40. Eds NC Stampolidis, A Kanta & A Giannikouri. University of Crete.

Belin BJ, Busset N, Giraud E, Molinaro A, Silipo A & Newman DK 2018 Hopanoid lipids: from membranes to plant-bacteria interactions. *Nature Reviews Microbiology* **16** 304-315.

Bloch K 1992 Sterol molecule: structure, biosynthesis, and function. *Steroids* **57** 378-383.

Bloch K & Rittenberg D 1942 On the utilization of acetic acid for cholesterol formation. *Journal of Biological Chemistry* **145** 625-636.

Bobrovskiy I, Hope JM, Ivantsov A, Nettersheim BJ, Hallmann C & Brocks JJ 2018 Ancient steroids establish the Ediacaran fossil Dickinsonia as one of the earliest animals. *Science* **361** 1246-1249.

Bridgham JT, Eick GN, Larroux C, Deshpande K, Harms MJ, Gauthier ME, Ortlund EA, Degnan BM & Thornton JW 2010 Protein evolution by molecular tinkering: diversification of the nuclear receptor superfamily from a ligand-dependent ancestor. *PLoS Biology* **8** e1000497.

Brocks JJ & Summons RE 2003 Sedimentary hydrocarbons, biomarkers for early life. *Treatise on Geochemistry* **8** 64-115.

Butterfield NJ 2015 Early evolution of the eukaryota. *Palaeontology* **58** 5-17.

Caforio A, Siliakus MF, Exterkate M, Jain S, Jumde VR, Andringa RLH, Kengen SWM, Minnaard AJ, Driessen AJM & van der Oost J 2018 Converting Escherichia coli into an archaebacterium with a hybrid heterochiral membrane. *Proceedings of the National Academy of Sciences U.S.A.* **115** 3704-3709.

Christianson DW 2017 Structural and chemical biology of terpenoid cyclases. *Chemical Reviews* **117** 11570-11648.

Cornforth JW, Cornforth RH, Popjak G & Gore IY 1958 Studies on the biosynthesis of cholesterol. 5. Biosynthesis of squalene from DL-3-hydroxy-3-methyl [2-14C] pentano-5-lactone. *Biochemical Journal* **69** 146-155.

Cronin JR & Pizzarello S 1990 Aliphatic hydrocarbons of the Murchison meteorite. *Geochimica et Cosmochimica Acta* **54** 2859-2868.

Cutler SR, Rodriguez PL, Finkelstein RR & Abrams SR 2010 Abscisic acid: emergence of a core signaling network. *Annual Review of Plant Biology* **61** 651-679.

Dagaut P & Cathonnet M 2006 The ignition, oxidation, and combustion of kerosene: a review of experimental and kinetic modelling. *Progress in Energy & Combustion Science* **32** 48-92.

Das S, Schapira M, Tomic-Canic M, Goyanka R, Cardozo T & Samuels HH 2007 Farnesyl pyrophosphate is a novel transcriptional activator for a subset of nuclear hormone receptors. *Molecular Endocrinology* **21** 2672-2686.

de Rosa M., Gambacorta A & Gliozzi A 1986 Structure, biosynthesis, and physicochemical properties of archaebacterial lipids. *Microbiological Reviews* **50** 70-80.

Desmond E & Gribaldo S 2009 Phylogenomics of sterol synthesis: insights into the origin, evolution, and diversity of a key eukaryotic feature. *Genome Biology and Evolution* **1** 364-381.

Dong LM, Wilson C, Wardell MR, Simmons T, Mahley RW, Weisgraber KH & Agard DA 1994 Human apolipoprotein E. Role of arginine 61 in mediating the lipoprotein preferences of the E3 and E4 isoforms. *Journal of Biological Chemistry* **269** 22358-22365.

Eme L, Spang A, Lombard J, Stairs CW & Ettema TJG 2017 Archaea and the origin of eukaryotes. *Nature Reviews Microbiology* **15** 711-723.

Eschenmoser A 1990 Leopold Ruzicka - from the isoprene rule to the question of life's origin. *Chimia* **44** 1-21.

Eschenmoser A & Arigoni D 2005 Revisited after 50 years: the stereochemical interpretation of the biogenetic isoprene rule for the triterpenes. *Helvetica Chimica Acta* **88** 3011-3150.

Eschenmoser A, Ruzicka L, Jeger O & Arigoni D 1955 Eine sterochemische interpretation der biogenetischen isoprenregl bei den triterpen. *Helvetica Chimica Acta* **38**1890-1904.

Euler W 1897 Ueber eine synthese und die constitution des isoprens. *Advances in Synthesis and Catalysis* **57** 131-159.

Faraday M 1826 On pure caoutchouc, and the substances by which it is accompanied in the state of sap, or juice. *Quarterly Journal of Science Literature & the Arts* **21** 19-28.

Galea AM & Brown AJ 2009 Special relationship between sterols and oxygen: were sterols an adaptation to aerobic life? *Free Radical Biology & Med.* **47** 880-889.

Garcia AK, Schopf JW, Yokobori SI, Akanuma S & Yamagishi A 2017 Reconstructed ancestral enzymes suggest long-term cooling of Earth's photic zone since the Archean. *Proceedings of the National Academy of Sciences U.S.A.* **114** 4619-4624.

Gelmont D, Stein RA & Mead JF 1981 Isoprene-the main hydrocarbon in human breath. *Biochemical & Biophysical Research Communications* **99** 1456-1460.

Gilmore SF, Yao AI, Tietel Z, Kind T, Facciotti MT & Parikh AN 2013 Role of squalene in the organization of monolayers derived from lipid extracts of Halobacterium salinarum. *Langmuir* **29** 7922-7930.

Goyanka R, Das S, Samuels HH & Cardozo T 2010 Nuclear receptor engineering based on novel structure activity relationships revealed by farnesyl pyrophosphate. *Protein Engineering Design & Selection* **23** 809-815.

Gribaldo S & Brochier-Armanet C 2006 The origin and evolution of Archaea: a state of the art. *Philosophical Transactions of the Royal Society London B Biological Sciences* **361** 1007-1022.

Gumsley AP, Chamberlain KR, Bleeker W, Soderlund U, de Kock MO, Larsson ER & Bekker A 2017 Timing and tempo of the Great Oxidation Event. *Proceedings of the National Academy of Sciences U.S.A.* **114** 1811-1816.

Haldane JBS 1929 Origin of life. *Rationalist Annual* **148** 3-10.

Ipatiew WI & Wittorf N 1897 Zur Constitution von Isopren. *Journal für Praktische Chemie* **55** 1-4.

Jia M & Peters RJ 2017 Cis or trans with class II diterpene cyclises. *Organic & Biomolecular Chemistry* **15** 3158-3160.

Jiang Z, Kempinski C & Chappell J 2016 Extraction and analysis of terpenes/terpenoids. *Current Protocols in Plant Biology* **1** 345-358.

Joyce GF, Schwartz AW, Miller SL & Orgel LE 1987 The case for an ancestral genetic system involving simple analogues of the nucleotides. *Proceedings of the National Academy of Sciences U.S.A.* **84** 4398-4402.

Langdon RG & Bloch K 1953 The biosynthesis of squalene. *Journal of Biological Chemistry* **200** 129-134.

Langworthy TA, Tornabene TG & Holzer G 1982 Lipids of archaebacteria. *Zentralblatt Fur Bakteriologie Mikrobiologie und Hygiene I Abteilung Originale C.* **3** 228-244.

Lathe R 2005 Tidal chain reaction and the origin of replicating biopolymers. *International Journal of Astrobiology* **4** 19-31.

Lazcano A, Oro J & Miller SL 1983 Primitive Earth environments: organic syntheses and the origin and early evolution of life. *Precambrian Research* **20** 259-282.

Liu XY & Sawant PD 2002 Micro/nanoengineering of the self-organized three-dimensional fibrous structure of functional materials. *Angewandte Chemie International Edition* **41** 3641-3645.

Lynen F, Eggerer H, Henning U & Kessel I 1958 Farnesyl Pyrophosphate und 3-Methyl-Ä^{3}-butenyl-1-pyrophosphat – Die biologischen Vorstufen des Squalens. *Angewandte Chemie.* **70** 738-742.

Lyons TW, Reinhard CT & Planavsky NJ 2014 The rise of oxygen in Earth's early ocean and atmosphere. *Nature.* **506** 307-315.

Martin WF, Garg S & Zimorski V 2015 Endosymbiotic theories for eukaryote origin. *Philosophical Transactions of the Royal Society London B Biological Sciences* **370** 20140330.

McGenity TJ, Crombie AT & Murrell JC 2018 Microbial recycling. *ISME Journal* **12** 931-941.

McGovern PE 2003 *Ancient Wine: The Search for the Origins of Viniculture*. Princeton University Press.

McGuire BA 2018 2018 Census of interstellar, circumstellar, extragalactic, protoplanetary disk, and exoplanetary molecules. *Astrophysical Journal Supplement Series* **239** 17.

Melendez I, Grice K & Schwark L 2013 Exceptional preservation of Palaeozoic steroids in a diagenetic continuum. *Scientific Reports* **3** 2768.

Miller SL & Urey HC 1959 Organic compound synthesis on the primitive earth. *Science* **130** 245-251.

Moore DD 1990 Diversity and unity in the nuclear hormone receptors: a terpenoid receptor superfamily. *Nature New Biology* **2** 100-105.

Nakatani Y, Ribeiro N, Streiff S, Desaubry L & Ourisson G 2012 Search for the most primitive membranes: some remaining problems. *Origins of Life and Evolution of Biospheres* **42** 497-501.

Nakatani Y, Ribeiro N, Streiff S, Gotoh M, Pozzi G, Desaubry L & Milon A 2014 Searth for the most 'primitive' membranes and their reinforcers: a review of the polyprenyl phosphates theory. *Origins of Life and Evolution of Biospheres* **44** 197-208.

Nes WD 2011 Biosynthesis of cholesterol and other sterols. *Chemical Reviews* **111** 6423-6451.

Oparin AI 1953 *The origin of life*. New York: Academic.

Oro J, Nooner DW, Zlatkis A & Wikstrom SA 1966 Paraffinic hydrocarbons in the Orgueil, Murray, Mokoia and other meteorites. *Life Sciences in Space Research* **4** 63-100.

Ourisson G 1989 The evolution of terpenes to sterols. *Pure and Applied Chemistry* **61** 345-348.

Ourisson G & Albrecht P 1992 Hopanoids. 1. Geohopanoids: the most abundant natural products on Earth? *Accounts of Chemical Research* **25** 398-402.

Ourisson G & Nakatani Y 1994 The terpenoid theory of the origin of cellular life: the evolution of terpenoids to cholesterol. *Chemistry & Biology* **1** 11-23.

Ourisson G, Rohmer M & Poralla K 1987 Prokaryotic hopanoids and other polyterpenoid sterol surrogates. *Annual Review of Microbiology* **41** 301-333.

Parker ET, Cleaves HJ, Dworkin JP, Glavin DP, Callahan M, Aubrey A, Lazcano A & Bada JL 2011 Primordial synthesis of amines and amino acids in a 1958 Miller H_2S-rich spark discharge experiment. *Proceedings of the National Academy of Sciences U.S.A.* **108** 5526-5531.

Pasek MA, Harnmeijer JP, Buick R, Gull M & Atlas Z 2013 Evidence for reactive reduced phosphorus species in the early Archean ocean. *Proceedings of the National Academy of Sciences U.S.A.* **110** 10089-10094.

Pastar I, Stojadinovic O, Sawaya AP, Stone RC, Lindley LE, Ojeh N, Vukelic S, Samuels HH & Tomic-Canic M 2016 Skin metabolite, farnesyl pyrophosphate, regulates epidermal response to inflammation, oxidative stress, and migration. *Journal of Cellular Physiology* **231** 2452-2463.

Pearson A, Budin M & Brocks JJ 2003 Phylogenetic and biochemical evidence for sterol synthesis in the bacterium Gemmata obscuriglobus. *Proceedings of the National Academy of Sciences U.S.A.* **100** 15352-15357.

Planavsky NJ, Asael D, Hofmann A, Teinhard CT, Lalonde DV, Knudsen A, Wang X, Ossa FA, Pecoits E, Smith AJB *et al.* 2014 Evidence for oxygenic photosynthesis half a billion years before the Great Oxidation Event. *Nature Geoscience* **7** 283-286.

Polke M, Leonhardt I, Kurzai O & Jacobsen ID 2018 Farnesol signalling in Candida albicans - more than just communication. *Critical Reviews in Microbiology* **44** 230-243.

Popa O, Babeanu NE, Popa I, Nita S & Dinu-Parvu CE 2015 Methods for obtaining and determination of squalene from natural sources. *Biomed Research International* **2015** 367202.

Pope EC, Bird DK & Rosing MT 2012 Isotope composition and volume of Earth's early oceans. *Proceedings of the National Academy of Sciences U.S.A.*. **109** 4371-4376.

Popjak G, Edmond J, Clifford K & Williams V 1969 Biosynthesis and structure of a new intermediate between farnesyl pyrophosphate and squalene. *Journal of Biological Chemistry* **244** 1897-1918.

Poralla K 2004 Profound insights into squalene cyclization. *Chemistry & Biology* **11** 12-14.

Prelog V & Jeger O 1980 Leopold Ruzicka. 13 September 1887-26 September 1976. *Biographical Memoirs of Fellows of the Royal Society* **26** 411-501.

Qu Z, Bendena WG, Tobe SS & Hui JHL 2018 Juvenile hormone and sesquiterpenoids in arthropods: Biosynthesis, signaling, and role of MicroRNA. *Journal of Steroid Biochemistry & Molecular Biology* **184** 69-76.

Reinert DJ, Balliano G & Schulz GE 2004 Conversion of squalene to the pentacarbocyclic hopene. *Chemistry & Biology* **11** 121-126.

Robinson R 1932 Constitution of cholesterol. *Nature* **130** 540-541.

Rohmer M, Bouvier P & Ourisson G 1979 Molecular evolution of biomembranes: structural equivalents and phylogenetic precursors of sterols. *Proceedings of the National Academy of Sciences U.S.A.* **76** 847-851.

Ruzicka L 1953 The isoprene rule and the biogenesis of terpenic compounds. *Experientia.* **9** 357-367.

Ruzicka L 1956 Bedeutung der theoretischen organischen Chemie für die Chemie der Terpenverbindungen. In *Perspectives in Organic Chemistry*, pp 265-314. Ed A Todd. Interscience Publishers.

Ruzicka L 1959 History of the isoprene rule. Faraday Lecture November 1959. *Proceedings of the Chemical Socety* **i** 341-376.

Ruzicka L 1963 Perspektiven der Biogenese und der Chemie der Terpene. *Pure and Applied Chemistry* **6** 493-524.

Ruzicka L 1966 Multimembered rings, higher terpene compounds and male sex hormones. Nobel Lecture, December 12, 1945. In *Nobel Lectures, Chemistry 1922–1941*, Elsevier.

Ruzicka L, Eschenmoser A, Heusser H 1953 Biogenesis of steroids and terpenic compounds. *Experientia* **9** p362.

Ruzicka T & Lardon F 1944 Über das Ambreïn, einen Bestandteil des grauen Ambra. *Helvetica Chimica Acta* **29** 912-921.

Saenz JP, Grosser D, Bradley AS, Lagny TJ, Lavrynenko O, Broda M & Simons K 2015 Hopanoids as functional analogues of cholesterol in bacterial membranes. *Proceedings of the National Academy of Sciences U.S.A.* **112** 11971-11976.

Schwartz AW 2006 Phosphorus in prebiotic chemistry *Philosophical Transactions of the Royal Society London B Biological Sciences* **361** 1743-1749.

Segré D, Ben-Eli D, Deamer DW & Lancet D 2001 The lipid world. *Origins of Life and Evolution of Biospheres* **31** 119-145.

Segré D & Lancet D 2000 Composing life. *EMBO Reports* **1** 217-222.

Sephton MA 2002 Organic compounds in carbonaceous meteorites. *Natural Product Reports* **19** 292-311.

Sharkey TD & Yeh S 2001 Isoprene emission from plants. *Annual Review of Plant Physiology and Plant Molecular Biology.* **52** 407-436.

Sleep NH 2010 The Hadean-Archaean environment. C*old Spring Harbor Perspectives in Biology* **2** a002527.

Spanova M, Zweytick D, Lohner K, Klug L, Leitner E, Hermetter A & Daum G 2012 Influence of squalene on lipid particle/droplet and membrane organization in the yeast Saccharomyces cerevisiae. *Biochimica et Biophysica Acta* **1821** 647-653.

Sprott GD 1992 Structures of archaebacterial membrane lipids. *Journal of Bioenergetics and Biomembranes* **24** 555-566.

Stonner C & Williams J 2016 European football: goals change crowd air chemistry. *Nature.* **535** 355.

Summons RE, Bradley AS, Jahnke LL & Waldbauer JR 2006 Steroids, triterpenoids and molecular oxygen. *Philosophical Transactions of the Royal Society London B Biological Sciences* **361** 951-968.

Summons RE, Powell TG & Boreham CJ 1988 Petroleum geology and geochemistry of the Middle Proterozoic McArthur Basin, northern Australia. III. Composition of extractable hydrocarbons. G*eochimica et Cosmochimica Acta* **52** 1747-1763.

Tavormina PA, Gibbs MH & Huff JW 1956 The utilization of beta-hydroxy-beta-methyl-delta-valerolactone in cholesterol biosynthesis. *Journal of the American Chemical Society* **78** 4498-4499.

Tilden WA 1884 On the decomposition of terpenes by heat. *Journal of the American Chemical Society* **45** 410-420.

Valitova JN, Sulkarnayeva AG & Minibayeva FV 2016 Plant sterols: diversity, biosynthesis, and physiological functions. *Biochemistry (Moscow)* **81** 819-834.

Waite JH, Jr., Young DT, Cravens TE, Coates AJ, Crary FJ, Magee B & Westlake J 2007 The process of tholin formation in Titan's upper atmosphere. *Science* **316** 870-875.

Wallach O 1887 Zur Kenntniss der Terpene und ätherischen Oele. *European Journal of Organic Chemistry* **238** 78-89.

Watson JD & Crick FH 1953 Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* **171** 737.

Wieland HO 1966 The chemistry of the bile acids. Nobel Lecture, December 12, 1928. In *Nobel Lectures, Chemistry 1922–1941*, pp 105-121.

Williams CG 1860 On isoprene and caoutchine. *Philosophical Transactions of the Royal Society London B Biological Sciences* **150** 241-255.

Wilson EH & Atreya SK 2018 Current state of modeling the photochemistry of Titan's mutually dependent atmosphere and ionosphere. *Journal of Geophysical Research* **109** E06002.

Woodward RG & Bloch K 1953 The cyclization of squalene in cholesterol synthesis. *Journal of the American Chemical Society* **75** 2023.

Wüersch J, Huang RL & Bloch K 1952 The origin of the isooctyl side chain of cholesterol *Journal of Biological Chemistry* **195** 439-446.

Xu R, Fazio GC & Matsuda SP 2004 On the origins of triterpenoid skeletal diversity. *Phytochemistry.* **65** 261-291.

Yamagata Y, Watanabe H, Saitoh M & Namba T 1991 Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature.* **352** 516-519.

Box 1. Isoprenoids and Terpenoids: History and Nomenclature

'Terpene' refers to a 10-carbon dimer of isoprene as well as to the generic class of chemical substances built of polyisoprene units. Although the IUPAC recommends that the terms 'terpenoid' and 'isoprenoid' are used to refer to chemically modified (generally oxygenated or methylated) forms of 'terpene'/'isoprene' polymers, these terms are all used interchangeably in the field. Confusingly, the molecule terpene, a 10-carbon isoprene dimer, is not itself the precursor for many terpenoids. Nevertheless, this nomenclature persists, and farnesol is regarded as a sesquiterpene (based on 3 isoprene units), whereas squalene, lanosterol, and steroids are triterpenoids (6 isoprene units). Because of their sterically constrained structures and multiple chiral forms, these volatile linear and (poly)cyclic terpenoids (i.e., 'aromatics') are well adapted as signaling molecules. Across the kingdom of life the terpene repertoire encompasses signaling molecules – both repellants/antimicrobials and attractants/pheromones – the latter being widely employed in perfumery. Ruzicka was keenly interested; he established that the fragrance of ambergris is based on the triterpene, ambrein (Ruzicka & Lardon 1944; Prelog & Jeger 1980). In the following we examine the origins of the key terms.

Terpene: named in 1863 by August Kekulé from turpentine/terpentine, the aromatic resin of the terebinth tree (*Pistacia terebinthus*) that grows widely around the Mediterranean – including Croatia, Ruzicka's birthplace. The major component of turpentine is a 10-carbon monoterpene (pinene). Tree resins protect the host tree against invasion by microbes and insects, and thus have potent medicinal properties – essential oils from terebinth (also 'tereminth') were highly prized and the tree had early religious significance (Barton 1906). A depiction of the terebinth leaf was a symbol in the Minoan writing system, and terebinth resins were used as a preservative for wine as early as 5400–5000 BCE (McGovern

2003). *Tereminthos* may be an early IndoEuropean word denoting/akin to 'overcomer of the forces of growth/death' (Beckmann 2012), with roots in *ter* (= *supra*) and *minth* (*cf* Greek *minth*, Latin *menta*, 'mint'; Greek *methu*, 'wine', also Welsh *medd*, English 'mead', and 'menthol', a monoterpenoid).

Isoprene: the 5-carbon isoprene molecule was (mis)named in 1860 by British chemist Charles H. Greville Williams as a supposed isomer of the propane/propyl group of substances, whose name ultimately derives from Greek piōn 'fat'. Any chemical substance formally derived from isoprene is an isoprenoid.

Farnesol: the key sesquiterpenoid is named from the floral essence of the acacia tree (*Vachellia farnesiana*) that was brought to Europe from the Americas by Cardinal Farnese (1573–1626).

Squalene: the triterpenoid (hexa-isoprene) squalene was first characterized in shark liver oil (Squalidae, the shark family) in 1906 by the Japanese scientist Mitsumaru Tsujimoto (Popa *et al.* 2015). Because squalene is lighter than water, cartilaginous fish (which lack a swim bladder) such as sharks reduce their body density with such fats/oils. However, squalene is widespread in plants and animals including human, where it is secreted by the liver, carried in the blood by LDL and VLDL, and secreted in large quantities from sebaceous glands where it may exert antimicrobial action (Popa *et al.* 2015).

Lanosterol: a triterpenoid formed by cyclization of squalene, the precursor to steroids, and a component of lanolin (from Latin *lāna* ‘wool’), the oily water-repellant secretion of sebaceous glands of sheep and other wool-producing animals. Derivatives have antimicrobial properties.

Of the primary counting system based on carbon numbers (meth-, eth-, prop-, but-; thereafter classical pent-, hex-, etc.), with the exception of 'eth-' (from *aether*, named in 1834/1835 by Justus Liebig and Jacob Berzelius), the others derive from 'wine', 'fat', and 'butter', respectively, with pride of place going to 'meth-', denoting wine, tereminth, and thus terpene.

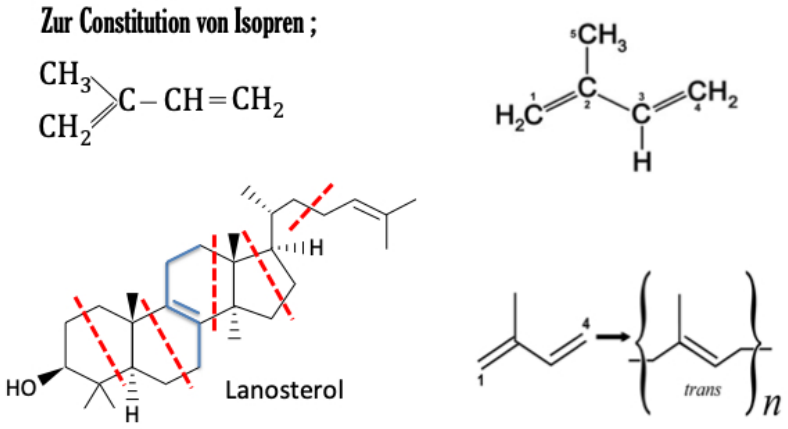

Figure 1. The Isoprene Unit. (Top left) Original depiction of the structure of isoprene (redrawn from Ipatiew & Wittorf (1897)), alongside (Top right) a contemporary rendition of the isoprene unit, with numbered carbon atoms, which is (Bottom right) the basic building block for polymeric isoprenoid synthesis through linear head-to-tail (C_1–C_4) condensation of individual isoprene units. (Bottom left) Diagram based on a sketch from Ruzicka's personal notes showing the tetracyclic structure of lanosterol broken (red dashed lines) into constituent 5-carbon isoprene units. The C_7–C_{12} isoprene unit at the boundary of rings B and C is highlighted in blue (redrawn from Eschenmoser (1990)).

254x190mm (72 x 72 DPI)

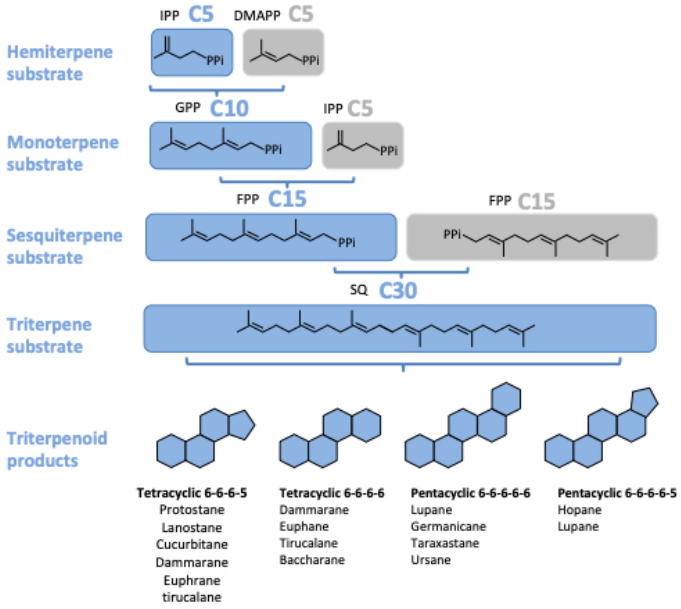

Figure 2. Simplified Scheme of Triterpenoid Formation by Stepwise Condensation of C_5 Isoprene Units (IPP, DMAPP) via C_{10} GPP and C_{15} FPP to Form C_{30} Squalene (SQ). Squalene is then cyclized by terpene/squalene cyclase enzymes into the multicyclic triterpenoid (carbon skeletons shown) substrates from which countless downstream secondary metabolites are formed, as discussed in the text. Abbreviations: DMAPP, dimethylallyl pyrophosphate (isomer of isoprene); FPP, farnesyl pyrophosphate; GPP, geranyl pyrophosphate; IPP, isopentenyl pyrophosphate.

254x190mm (72 x 72 DPI)

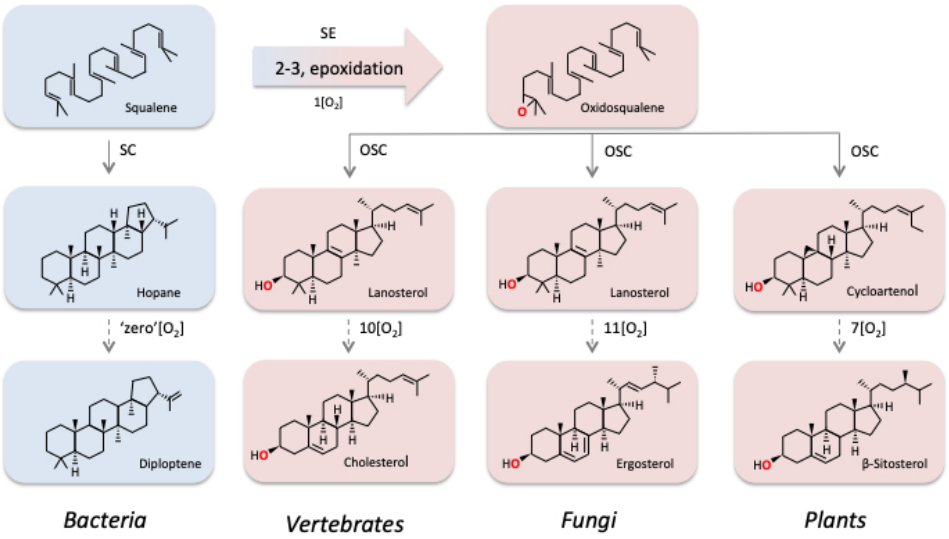

Figure 3. Overview of Squalene Cyclization during Sterol Biosynthesis. In bacteria, cyclization of squalene catalyzed by squalene cyclase (SC) does not require oxygen (blue buttons). In plants, animals, and fungi (pink buttons), squalene oxidation catalyzed by squalene epoxidase (SE) is required before cyclization catalyzed by oxidosqualene cyclase (OSC) begins. The minimum number of O_2 molecules required to convert one molecule of squalene to one molecule of sterol is calculated to be 0, 8, 11, and 12, for diploptene, β-sitosterol, cholesterol, and ergosterol, respectively. Data from Summons *et al.* (2006).

254x190mm (72 x 72 DPI)

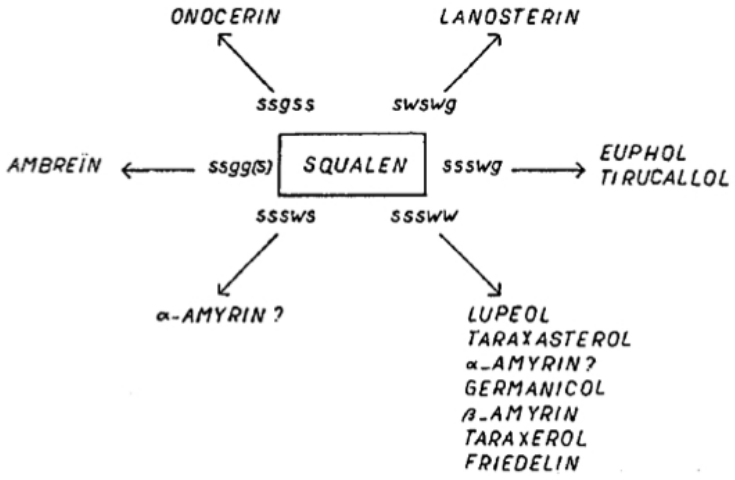

Figure 4. The Squalene Hypothesis of Triterpenoid Biosynthesis. Summary of the major squalene cyclization events, as understood by Eschenmoser *et al.* (1955), 10 years after the delivery of Ruzicka's Nobel lecture. Abbreviations: *s*, chair folding; *w*, boat folding; *g*, stretched conformation of the squalene chain. Reproduced, with permission, from Eschenmoser & Arigoni (2005).

254x190mm (72 x 72 DPI)

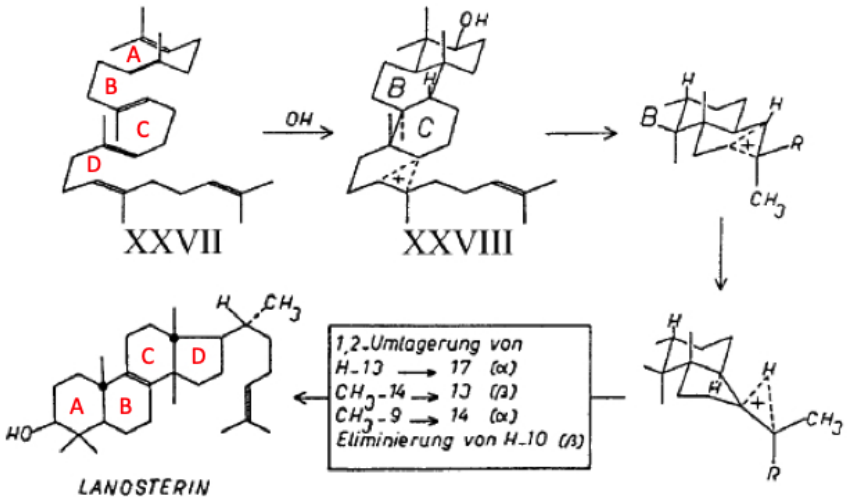

Figure 5. Molecular Mechanism of Squalene (**XXVII**) Cyclization to Lanosterol (*Lanosterin*) via Oxidosqualene (**XXVIII**) as Originally Proposed by Ruzicka's Team (Eschenmoser *et al.* 1955). The tetracyclic triterpenoid footprint is highlighted (red lettering). Reproduced, with permission, from Eschenmoser & Arigoni (2005).

254x190mm (72 x 72 DPI)

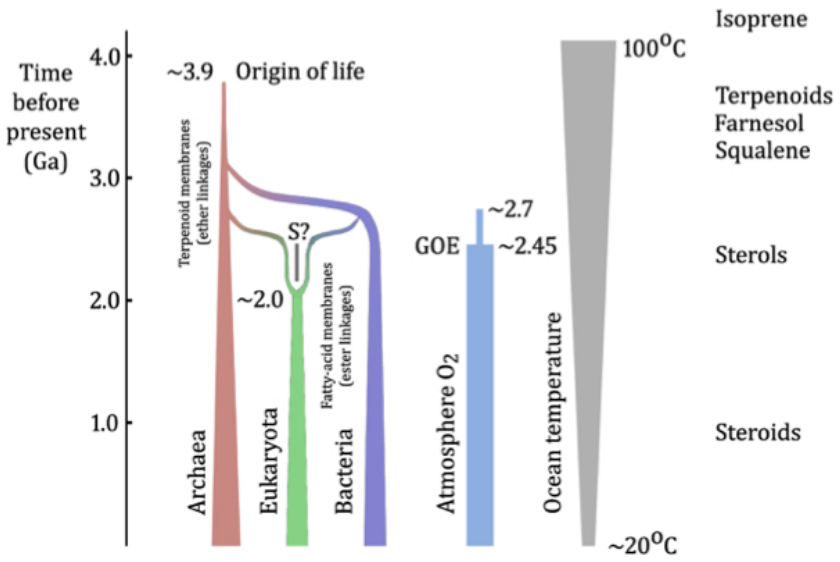

Figure 6. Terpenoids and the Diversification of Life Forms. S indicates a period of possible symbiosis between Archaea and Bacteria that may have preceded the generation of the first eukaryote generated by incorporation of an alphaproteobacterium within an archaeal host cell. Ga, billion years; GOE, Great Oxidation Event.

254x190mm (72 x 72 DPI)

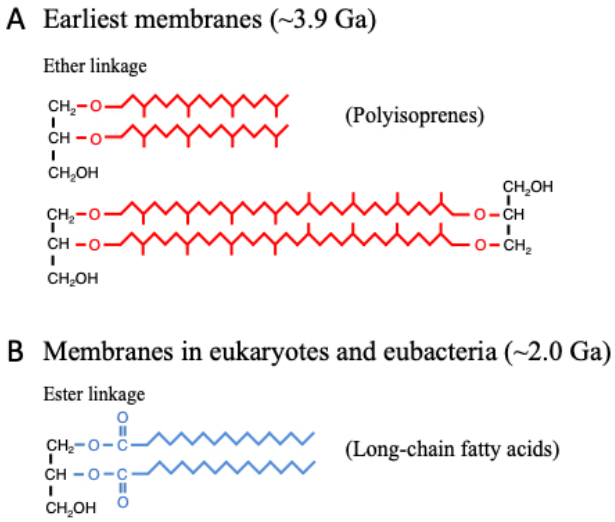

Figure 7. Long-Chain hydrocarbons in Archaeal and Modern Membranes. (A) Typical membrane lipids in Archaea, comprising long-chain terpenoids attached to glycerol via ether linkages. (B) Typical lipids in eukaryotes and eubacteria comprising long-chain fatty acids linked to glycerol via ester bonds. Figure redrawn from Langworthy *et al.* (1982).

254x190mm (72 x 72 DPI)